Discrete Mathematics 75 (1989) 23–30
North-Holland

GRAPHS WITH A SMALL NUMBER OF DISTINCT INDUCED SUBGRAPHS

Noga ALON*

Department of Mathematics, Sackler Faculty on Exact Sciences, Tel Aviv University, Tel Aviv, Israel

Béla BOLLOBÁS†

Department of Mathematics, Cambridge University, Cambridge CB2 1SB, England

Let G be a graph on n vertices. We show that if the total number of isomorphism types of induced subgraphs of G is at most εn^2, where $\varepsilon < 10^{-21}$, then either G or its complement contain an independent set on at least $(1-4\varepsilon)n$ vertices. This settles a problem of Erdős and Hajnal.

1. Introduction

All graphs considered here are finite, simple and undirected. For a graph G, let $i(G)$ denote the total number of isomorphism types of induced subgraphs of G. We call $i(G)$ the *isomorphism number* of G. Note that $i(G) = i(\bar{G})$, where $\bar{G}$ is the complement of G, and that if G has n vertices then $i(G) \geq n$, as G contains an induced subgraph with m vertices for each m, $1 \leq m \leq n$. An induced subgraph H of G is called *trivial* if it is either complete or independent. Let $t(G)$ denote the maximum number of vertices of a trivial subgraph of G. Note that the complete bipartite graph G with vertex classes of size $n/2$ (>1) each has $t(G) = n/2$ and $i(G) = \Theta(n^2)$. The above two estimates hold for a matching of $n/2$ edges, too. In March 1988, at the Cambridge Combinatorial Conference, András Hajnal conjectured that if G is a graph on n vertices and $i(G) = o(n^2)$, then $t(G) = n - o(n)$. As the main result of this paper, we shall prove this conjecture. Independently of us, the conjecture was proved in a stronger form by Erdős and Hajnal [2].

Theorem 1.1. *Let G be a graph on n vertices. If $i(G) \leq \varepsilon n^2$, where $\varepsilon < 10^{-21}$, then $t(G) \geq (1-4\varepsilon)n$.*

It is worth noting that both constants 10^{-21} and 4 in the theorem above can be improved easily. We make no attempt to optimize the constants here and in the rest of the paper.

The proof of Theorem 1.1 is somewhat lengthy, and is presented in the next

* Research supported in part by Allon Fellowship, by a Bat-Sheva de Rothschild grant and by the Fund for Basic Research Administered by the Israel Academy of Sciences.
† Research supported in part by NSF Grant DMS 8806097.

three sections. We first consider, in Section 4, graphs G which contain relatively large trivial subgraphs. Somewhat paradoxically, graphs with no large trivial subgraphs are more difficult to deal with; this case with be discussed in Section 3. In Section 1, Theorem 1.1 is obtained as an easy consequence of the results of Sections 2 and 3. In the final section we present some unsolved problems.

2. Graphs with large trivial subgraphs

In this section we prove the following theorem, which implies the assertion of Theorem 1.1 for graphs with relatively large trivial subgraphs.

Theorem 2.1. *Let G be a graph on n vertices and put $t = t(G)$. Then*

$$i(G) \geq \min\left\{\frac{t^2}{3}, \frac{t(n-t)}{3}\right\}.$$

This theorem is an easy consequence of the following lemma.

Lemma 2.2. *Let $G = (V, E)$ be a graph on n vertices and put $t = t(G)$. If $t \geq n/2$ then*

$$i(G) \geq t(n-t)/3.$$

Proof. By replacing, if necessary, G by its complement, we may assume that there is an independent set T of t vertices. Let H be the bipartite subgraph of G with vertex classes T and $V \setminus T$ whose edges are the edges of G joining a vertex of T to a vertex of $V \setminus T$. Let $M = \{a_1b_1, a_2b_2, \dots, a_sb_s\}$ be a maximal matching in H, where $a_1, \dots, a_s \in T$ and $b_1, \dots, b_s \in V \setminus T$. Furthermore, set $A = \{a_1, \dots, a_s)$, $B = \{b_1, \dots, b_s\}$, $C = V \setminus (T \cup B)$ and $r = |C| = n - t - s$. Note that by the maximality of M there are no edges from C to $T \setminus A$.

Given l and m satisfying $0 \leq l \leq s$, $l \leq m \leq t$ and $m \geq 1$, let T' be a subset of $T \setminus \{a_1, a_2, \dots, a_l\}$ of cardinality $m - l$ and let $G_{l,m}$ be the subgraph of G spanned by the set of vertices $\{a_1, b_1, a_2, b_2, \dots, a_l, b_l\} \cup T'$. It is easily checked that $G_{l,m}$ has $l + m$ vertices and that its independence number is m. Therefore no two distinct members of the family $\{G_{l,m} : 0 \leq l \leq s, l \leq m \leq t, m \geq 1\}$ are isomorphic and hence

$$i(G) \geq t + t + (t-1) + (t-2) + \cdots + (t-s+1) = t(s+1) - \binom{s}{2}. \qquad (1)$$

Similarly, for each p, $0 \leq p \leq r$ and each q, $0 \leq q \leq t - s$, let $H_{p,q}$ be the induced subgraph of G on $C' \cup A \cup T'$, where $C' \subset C$ is a subset of C of cardinality p, and $T' \subset T \setminus A$ is a subset of T with $|T'| = q$. Since in G there are no edges from C to $T \setminus A$ it is easy to check that $H_{p,q}$ has $p + q + s$ (≥ 1) vertices, and that its independence number is $q + s$. Thus

$$i(G) \geq (r+1)(t-s+1) = (n-t-s+1)(t-s+1). \tag{2}$$

We shall make no attempt to obtain the best bound implied by inequalities (1) and (2); we shall prove only the claim of the lemma.

Multiplying inequality (1) by two and adding to it inequality (2), we see that

$$\begin{aligned} 3i(G) &\geq 2t(s+1) - s(s-1) + (n-t+1-s)(t+1-s) \\ &= (2t-n)s + 2t - s + (n-t+1)(t+1) \geq (n-t+1)(t+1). \quad \square \end{aligned}$$

Proof of Theorem 2.1. Let $G = (V, E)$ be a graph on n vertices and put $t = t(G)$. If $t \geq n/2$ then the assertion of the theorem follows from Lemma 2.2. Otherwise, let $T \subset V$ be the set of vertices of a trivial subgraph of G, with $|T| = t$. Let U be an arbitrary subset of cardinality t of $V \setminus T$, and let H be the induced subgraph of G on $T \cup U$. Clearly $t(H) = t = \frac{1}{2}|T \cup U|$ and hence, by Lemma 2.2 $i(G) \geq i(H) \geq t^2/3$. This completes the proof. $\square$

3. Graphs without large trivial subgraphs

This section is the heart of the paper; our main aim is to prove the following result.

Theorem 3.1. *Let G be a graph on n vertices. If $t(G) < n/10^{10}$ then $i(G) \geq n^2/10^{10}$.*

The proof of this result is rather long and is based on two propositions. In turn, in the proofs of these propositions we make use of the following very useful lemma of Erdős and Lovász [3] (see also [1, pp. 20–22]) sometimes called the Erdős–Lovász Local Lemma.

Lemma 3.2. *Let $A_1, \ldots, A_s$ be events in a probability space and let H be a graph of maximal degree $d \geq 2$ on the set $\{1, 2, \ldots, s\}$. Suppose that each A_i is independent of the system $\{A_j$: i is not joined to j in $H\}$ and $P(A_i) < 1/ed$. Then the probability that no A_i occurs is positive.*

Proposition 3.3. *Let $G = (V, E)$ be a graph of order n and maximal degree Δ, with $10^8 \leq \Delta \leq 0.9n$. Then for every two integers j and l that satisfy*

$$0.51\Delta < j < 0.52\Delta < l \leq 0.5n$$

there is an induced subgraph H of G with $|V(H)| = l$ vertices and maximal degree $\Delta(H) = j$. In particular $i(G) \geq n\Delta/10^4$.

Proof. Let $f: V \to \{0, 1\}$ be a random function, i.e. a random two-colouring of V obtained by choosing, for each $u \in V$ independently, a colour $f(u) \in \{0, 1\}$ according to a uniform distribution on $\{0, 1\}$. For each vertex $u \in V$, let A_u be the event that u has more than $(\Delta/2) + 2\sqrt{\Delta \log \Delta}$ neighbours having the same colour. By the standard estimates for the probability in the tail of the binomial distribution (see e.g. [1. p. 13, Theorem 7]), it is easy to check that for every $u \in V$ we have

$$P(A_u) \leq \Delta^{-8/3}.$$

Let H be the square of G, i.e. the graph obrained from G by adding all edges joining vertices at distance 2. Then $\Delta(H) \leq \Delta(\Delta - 1) < \Delta^{8/3}/e$, and so the graph H and the events A_u, $u \in V$, satisfy the conditions of Lemma 3.1. Therefore, with positive probability no A_u occurs.

Since $(\Delta/2) + 2\sqrt{\Delta \log \Delta} < 0.51\Delta - 3$, there is a two-colouring $f: V \to \{0, 1\}$ in which no vertex has more than $0.51\Delta - 3$ neighbours of either colour. We may assume without loss of generality that f gives colour 0 to at least half of the vertices: $|f^{-1}(0)| \geq n/2$. Set $U = f^{-1}(0) \cup \{v\}$, where v is a vertex of maximal degree Δ in G. Note that no vertex of H has more than $0.51\Delta - 2$ neighbours in U.

We next construct a sequence $H_0, H_1, \ldots, H_r$ of induced subgraphs of G with the following four properties:

(a) $\Delta(H_0) = \Delta$,

(b) $U \subset V(H_i)$ for every i,

(c) $\Delta(H_i) - 1 \leq \Delta(H_{i+1}) \leq \Delta(H_i)$ for every i,

(d) $\Delta(H_r) \leq 0.51\Delta$.

To construct this sequence we start by taking $H_0 = G$. Suppose that $H_0, H_1, \ldots, H_p$ have already been defined and they satisfy (a), (b) and (c). If $\Delta(H_p) \leq 0.51\Delta$ we take $p = r$ and complete the construction. Otherwise, H_p has at least 2 vertices that do not belong to U. If one of them is a vertex of maximal degree in H_p, we obtain H_{p+1} by deleting the other. Otherwise, let H_{p+1} be the graph obtained from H_p by deleting one of these vertices. One can easily check that $H_0, H_1, \ldots, H_{p+1}$ satisfy (a), (b) and (c) and hence we can continue this process and complete the construction.

By property (b) each H_i has at least $n/2$ vertices. By properties (a), (c) and (d) for each j, $0.51\Delta \leq j \leq \Delta$, one of these graphs has a maximal degree j. By deleting from such a graph all the non-neighbours of a vertex of maximal degree, one by one, we conclude that for every l satisfying $j \leq l \leq n/2$, there is an induced subgraph of G with l vertices and maximal degree j. In particular, there is a family of graph satisfying the conclusion of Proposition 3.3. □

The following technical result is a more complicated variant of the previous proposition.

Proposition 3.4. *Let $G = (V, E)$ be a graph on n vertices with maximal degree $\Delta < n/100$. Suppose furthermore that the independence number of G is at most $n/10^8$. Then, for every two integers j and l satisfying*

$$0.51\Delta < j < \Delta, \quad 0.05n \leq l \leq 0.49n$$

there is an induced subgraph $H = (V(H), E(H))$ of G, with no isolated vertices and with maximal degree $\Delta(H)$ such that

$$j - 1 \leq \Delta(H) \leq j \quad \text{and} \quad l \leq |V(H)| \leq l + 1.$$

In particular, G contains more than $\Delta n/100$ induced, pairwise non-isomorphic subgraphs, with no isolated vertices.

Proof. The proof is similar to the previous one but contains several additional complications.

Let $v \in V$ be a vertex of maximal degree $d(v) = \Delta$ in G, and denote by $N(v) = \Gamma(v) \cup \{v\}$ the set of neighbours of v together with the vertex v. Let $M = \{a_1b_1, a_2b_2, \ldots, a_sb_s\}$ be a maximum matching in the induced subgraph of G on $V - N(v)$. Put $U = N(v) \cup \{a_1, b_1, a_2, b_2, \ldots, a_s, b_s\}$ and let H be the induced subgraph of G on U. By the maximality of M, $V \setminus U$ is an independent set in G and hence

$$|U| \leq n - \frac{n}{10^8}. \tag{3}$$

Let $f: U \to \{0, 1\}$ be a random two-colouring of U obtained as follows: for each $u \in N(v) \cup \{a_1, a_2, a_3, \ldots, a_s\}$, the colour $f(u) \in \{0, 1\}$ of u is chosen according to a uniform distribution on $\{0, 1\}$ with all choices being independent. For all $1 \leq i \leq s$ define $f(b_i) = f(a_i)$. For each vertex $u \in U$ let A_u be the event that u has more than $(\Delta/2) + 3\sqrt{\Delta \log \Delta}$ neighbours in H having the same colour. As before, standard estimates for the binomial distribution (see [1, p. 13, Theorem 7]) imply that for every $u \in U$ we have

$$P(A_u) < \Delta^{-6}.$$

Clearly, each event A_u is independent of the system of events $\{A_w : w \in U, d(u, w) \geq 5\}$. Since for $u \in U$ at most $2\Delta^4$ events A_w, $w \in U$, do not belong to this system, and $P(A_u) < \Delta^{-6} < (2\Delta^4 e)^{-1}$, by Lemma 3.2 the probability that no event A_u occurs is positive.

Since $(\Delta/2) + 3\sqrt{\Delta \log \Delta} < 0.51\Delta - 5$, there is at least one two-colouring f of U in which no vertex has more than $0.51\Delta - 5$ neighbours in H having the same colour. Without loss of generality we may assume that there is a set U_1' of at least $|U|/2$ vertices of U all coloured 0. Put $U_1 = U_1' \cup \{v\}$. Note that no vertex of H has more than $0.15\Delta - 4$ neighbours in U_1.

Next, we construct a sequence $H_0, H_1, \ldots, H_r$ of induced subgraphs of H with the following five properties:

(a) $\Delta(H_0) = \Delta$,

(b) $U_1 \subset V(H_i)$ for every i,

(c) for each H_i and each $1 \leq j \leq s$ the vertex a_j belongs to H_i iff b_j is a vertex of H_i,

(d) $\Delta(H_i) - 2 \leq \Delta(H_{i+1}) \leq \Delta(H_i)$ for every i,

(e) $\Delta(H_r) \leq 0{\cdot}51\Delta$.

To construct such a sequence we start by taking $H_0 = H$. Suppose that $H_0, H_1, \ldots, H_p$ have already been defined and they satisfy (a), (b), (c) and (d). If $\Delta(H_p) \leq 0.51\Delta$ we take $p = r$ and complete the construction. Otherwise the graph H_p has at least four vertices that do not belong to U_1. We construct H_{p+1} by deleting one or two vertices of H_p as follows. Let u_0 and u_1 be two of these four vertices such that u_0u_1 is not an edge of M. If one of these vertices u_i is a vertex of maximal degree in H_p, we obtain H_{p+1} by deleting the other vertex u_{1-i} and the vertex to which u_{1-i} is matched under M, if there is such a vertex. Otherwise, we obtain H_{p+1} by deleting u_0 and the vertex to which it is matched under M, if there is such a vertex. It is easily checked that the sequence $H_0, H_1, H_2, \ldots, H_{p+1}$ also satisfies (a), (b), (c) and (d). Since H_{p+1} has fewer vertices than H_p, the process will end within n steps.

It is obvious that none of the graphs $H_0, H_1, \ldots, H_r$ has isolated vertices, and by (3) each of them has more than $0.49n$ vertices. Moreover, properties (a), (d) and (e) imply that for each j, $0.51\Delta < j \leq \Delta$, at least one of these graphs has maximal degree j or $j-1$. Let z_i be a vertex of maximal degree (j or $j-1$) in such a graph H_i. Since $\Delta \leq n/100$, z_i and its neighbours in H_i are incident with at most $n/100 + 1$ edges of M that saturate less than $0.05n$ vertices of H_i. By successively deleting all the non-neighbours of z_i in H_i, in such a way that together with every vertex matched under M we delete its mate as well, we conclude that for every $0.05n \leq l \leq |V(H_i)|$ there is an induced subgraph of H_i without isolated vertices of H_i, with either l or $l+1$ vertices and with maximal degree $\Delta(H_i) \in \{j-1, j\}$. This completes the proof of Proposition 3.4. □

Proof of Theorem 3.1. Let $G = (V, E)$ be a graph on n vertices satisfying $t(G) \leq n/10^{10}$. By replacing, if necessary, G by its complement, we may assume that $|E| \leq \frac{1}{2}\binom{n}{2}$. This easily implies the existence of an induced subgraph $H = (V, E)$ of G on $m \geq n/10$ vertices with maximal degree $\Delta \leq 0.9m$. Indeed, otherwise there is a sequence $v_1, v_2, \ldots, v_{\lfloor 0{\cdot}9n \rfloor}$ of vertices of G so that v_i has degree greater than $0.9(n-i)$ in the induced subgraph of G on $V(G)\backslash \{v_1, \ldots, v_{i-1}\}$. But in this case $|E(G)| \geq 0.9n + 0.9(n-1) + \cdots + 0.9(n - \lfloor 0.9n \rfloor) > \frac{1}{2}\binom{n}{2}$, a contradiction.

Clearly $t(H) \leq t(G) < m/10^8$. Let $\Delta = \Delta(H)$ denote the maximal degree in H. Since the independence number of H is smaller than $m/10^8$, we have $\Delta > 10^8 - 1$. If $\Delta \geq m/1000$ then, by Proposition 3.3, $i(G) \geq i(H) \geq m\Delta/10^4 \geq m^2/10^7 \geq$

$n^2/10^8$, implying the assertion of Theorem 3.1. Thus we may assume that

$$10^8 \leqslant \Delta = \Delta(H) \leqslant m/1000. \tag{4}$$

Let v be a vertex of maximal degree in H and let $\Gamma_H(v) = \{v_1, v_2, \dots, v_\Delta\}$ be the set of all its neighbours. Clearly $\sum_{u \in V(H)} |\Gamma_H(u) \cap \Gamma_H(v)| = \sum_{i=1}^{\Delta} d_H(v_i) \leqslant \Delta^2$, and hence the number of vertices $u \in V(H)$ for which $|\Gamma_H(u) \cap \Gamma_H(v)| > 10\Delta^2/m$ does not exceed $m/10$. Let us call a vertex $u \in V(H)$ *good* if $u \neq v$, u is not a neighbour of v and $|\Gamma_H(u) \cap \Gamma_H(v)| \leqslant 10\Delta^2/m$. Clearly, the number of good vertices in H is at least $m - \Delta - (m/10) > m/2$. We now construct a set $\{u_1, u_2, \dots, u_r\}$, with $r = \lceil m/100\Delta \rceil$, as follows. Let u_1 be a good vertex of H and put $H_1 = H \setminus N_H(u_1)$ where, as earlier, $N_H(u) = \{u\} \cup \Gamma_H(u)$. Clearly, H_1 has at least $m - (\Delta + 1) > m/2$ vertices and thus it has at least one good vertex. Let u_2 be such a vertex and put $H_2 = H_1 \setminus N_{H_1}(u_2)$. This process can be continued for at least r steps, since after $l \leqslant r$ steps we are still left with at least $m - l(\Delta + 1) \geqslant m - \lceil m/100\Delta \rceil (\Delta + 1) > m/2$ vertices. Note that the degree of v in H_r is at least $\Delta - (10\Delta^2/m) \cdot \lceil m/100\Delta \rceil > \Delta/2$, that $\{u_1, \dots, u_r\}$ is an independent set of vertices in G and that no u_i has a neighbour in H_r. The graph H_r has $m' \geqslant m/2$ vertices and maximal degree Δ' satisfying $\Delta/2 \leqslant \Delta' \leqslant \Delta$. Moreover, $t(H_r) \leqslant t(G) < n/10^{10} < m/10^8$. By inequality (4), we have $\Delta' \leqslant m'/100$. Therefore, by Proposition 3.4, the graph H_r contains at least $\Delta' m'/100 \geqslant \Delta m/200$ induced, pairwise non-isomorphic subgraphs, with no isolated vertices. All the induced subgraphs of G obtained by taking one of these subgraphs together with a set $\{u_1, u_2, \dots, u_s\}$, $0 \leqslant s \leqslant r$, are pairwise non-isomorphic, since $u_1, \dots, u_s$ are the only isolated vertices in each of these subgraphs. We thus conclude that

$$i(G) \geqslant r \cdot (m\Delta/200) \geqslant (m/100\Delta) \cdot (m\Delta/200) \geqslant m^2/10^5 \geqslant n^2/10^7.$$

This completes the proof of Theorem 3.1. $\square$

4. The proof of the main result

In this short section, we finally deduce Theorem 1.1 from the results of the previous two sections. Let G be a graph on n vertices, and suppose that $i(G) \leqslant \varepsilon n^2$, where $\varepsilon < 10^{-21}$. By Theorem 3.1 we have $t(G) \geqslant n/10^{10}$. Put $t(G) = t$. By Theorem 2.1 we have $t \geqslant n/2$ since otherwise $i(G) \geqslant t^2/3 \geqslant n^2/3.10^{20} > n^2/10^{21}$ contradicting the hypothesis. Therefore, by Theorem 2.1, $t(n-t)/3 < \varepsilon n^2$. Since $t \geqslant n/2$ and $\varepsilon < 10^{-21}$, this easily gives $t \geqslant (1 - 4\varepsilon)n$, completing the proof of Theorem 1.1. $\square$

5. Unsolved problems

In proving our theorem, we did not count the total number of isomorphism types of induced subgraphs, as the definition of $i(G)$ requires, but only the total number of types that can be distinguished by the following five parameters: the order, the maximal degree, the independence number, the clique number and the number of isolated vertices. In fact, in any particular case, we used only two of these parameters to show that we had sufficiently many non-isomorphic subgraphs. This raises the following rather general question: given a set Π of graph parameters and a graph G of order n with $t = t(G)$, at least how many isomorphism classes of induced subgraphs are there in G that can be distinguished by the parameters in Π? Writing $f(n, t; \Pi)$ for the minimum, our main result shows that if $\varepsilon > 0$ is small enough then for $t \leqslant (1-\varepsilon)n$ we have $f(n, t; \Pi_0) \geqslant \varepsilon n^2/4$, where Π_0 is the set of five parameters above. It would be interesting to determine, whether a similar inequality is true for the set Π, consisting of order and size.

In fact, the following more general problem presents itself. Given a set $\mathcal{H}_n$ of graphs of order n, and a set Π of graph parameters, what is the minimum of the number of induced subgraphs in a graph $H \in \mathcal{H}_n$ distinguished by Π? In this paper we studied the set of graphs without large trivial subgraphs.

One could also hope for considerably sharper results concerning the connection between $i(G)$ and $t(G)$. Is it true for every $\varepsilon > 0$ and natural number k, there is a constant $c = c(\varepsilon, k) > 0$ such that if G_n is a graph of order n satisfying $t(G_n) \leqslant (1/k - \varepsilon)n$ then $i(G) \geqslant cn^{k+1}$? At the moment we cannot even show that if $t(G_n) = o(n)$ then $i(G_n)$ grows faster than any polynomial of n.

Finally, let us state one of the problems of Erdős and Rényi: given $c > 0$, is there a constant $d = d(c) > 0$ such that if $t(G_n) \leqslant c \log n$ then $i(G) \geqslant 2^{dn}$?

Acknowledgement

This research was done while we were visiting the I.H.E.S. in Bures Sur Yvette, France, We would like to thank our hosts in I.H.E.S. for their hospitality. We are also indebted to Gil Kalai for stimulating discussions.

References

[1] B. Bollobás, Random Graphs (Academic Press, London, 1985) xvi + 447pp.
[2] P. Erdős and A. Hajnal, On the number of distinct induced subgraphs of a graph, this issue.
[3] P. Erdős and L. Lovász, Problems and results on 3-chromatic hypergraphs and some related questions, in 'Infinite and Finite Sets', A. Hajnal et al., eds. (North-Holland, Amsterdam, 1975) 609–628.